Data Assimilation and its applications

TNO | Knowledge for business

Delft University of Technology

Inverse Problem – Conceptual understanding

- The forward problem can be conceptually formulated as follows:

  Model parameters → Data

- The inverse problem - relates the model parameters to the data that we observe:

  Data → Model parameters

- The transformation from data to model parameters (or vice versa) is a result of the interaction of a physical system with the object that we wish to infer properties about.

Some examples

Physical system	**Governing equations**	**Physical quantity**	**Observed data**
Earth's gravitational field	Newton's law of gravity	Density	Gravitational field
Earth's magnetic field (at the surface)	Maxwell's equations	Magnetic suceptibility	Magnetic field
Seismic waves (from earthquakes)	Wave equation	Wave-speed (density)	Particle velocity

Key elements for successful solution?
Cooking book

APPETITE

COOKING SKILLS

TYPE OF EATERS

PASSION / CHEMISTRY

WILLINGNESS TO MIX & EXPERIMENT

NEED / CHALLENGE

'FUNDAMENTAL' KNOWLEDGE

DOMAIN KNOWLEDGE

PASSIONATE / CREATIVE PEOPLE

CROSS FERTILIZATION

List of recipes

- Optimal interpolation
- Kriging
- Variational methods
- Ensemble methods
- Hybrid methods

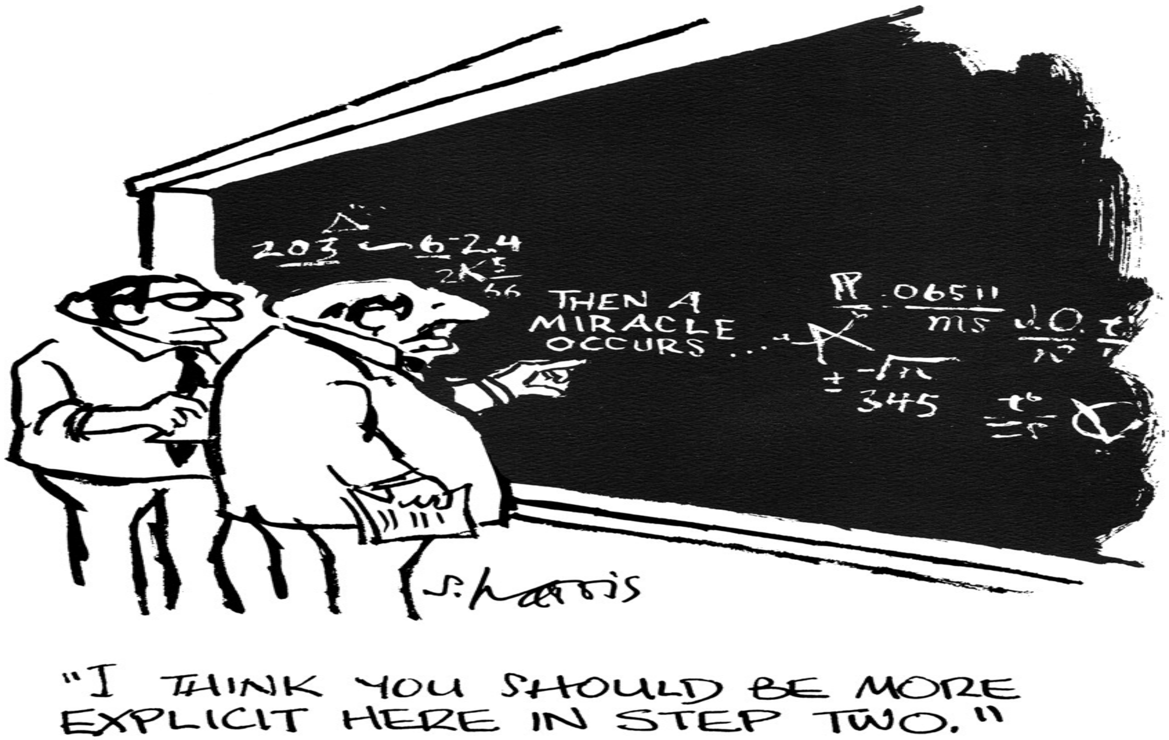
Data Assimilation – Main ingredients

- Two sources of information about the true state of the nature:
  - **Model** (abstraction of reality in terms of a set of differential equations)
  - **Measurements** (measure of certain quantities of interest)
- **Uncertainties** are present in both worlds.
- **Prior knowledge** (expert opinion)

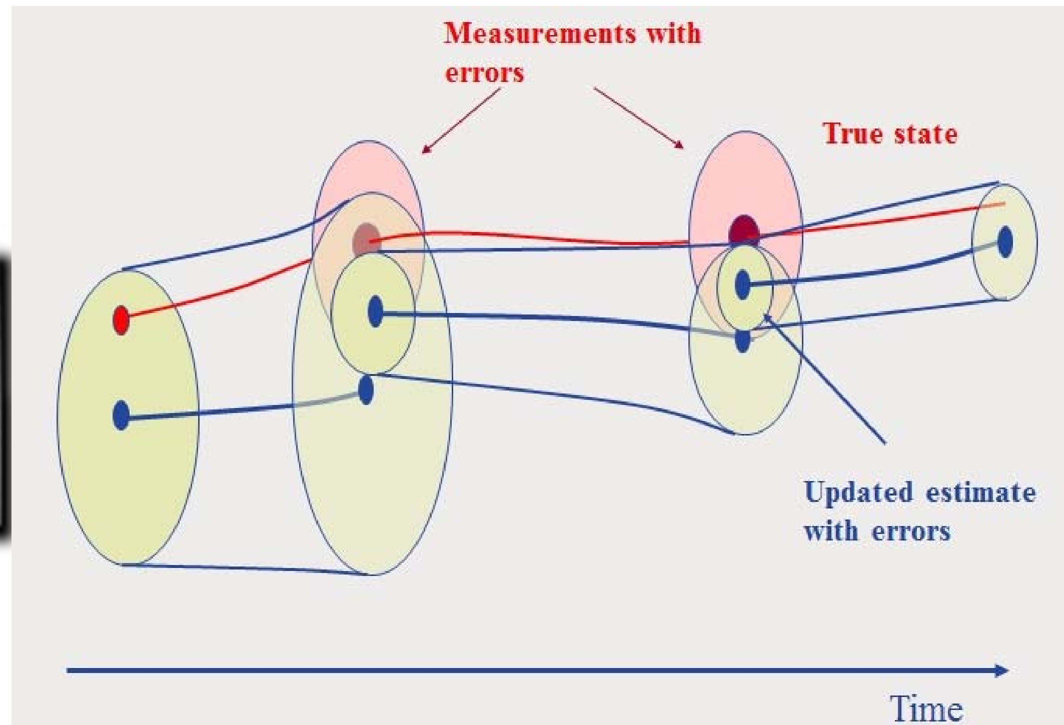
TU Delft
Delft University of Technology

TNO

- **The goal**: **An optimal estimate** of the truth based on the combination of both uncertain sources of information



The goal: **An optimal estimate** of the truth based on the combination of both uncertain sources of information



A different flavour for everyone / every challenge

Did you notice how a country-specific cuisine tasted differently in said country and abroad?

- Chinese food tastes like Indonesian in Netherlands and like Vietnamese in France .

- Italian pizza you have at your local Italian restaurant is rarely the same as the one you have in Italy.

Foods are tailored to meet the specific preferences of each country

A different flavour for everyone / every challenge

"The podiatrist wants jam on his toast, the psychiatrist wants nuts on his cereal, the plastic surgeon wants no wrinkles on her bacon, and the fertility doctor wants his eggs frozen."

Model and its uncertainties

Petroleum Geosciences

- Geological uncertainties
- Un-modelled physics
- Different scales
- ...

Fluid Dynamics

- Grain dimensions
- Un-modelled physics

Climate, Air and Sustainability

- Unknown sources
- Reaction rates
- Un-modelled physics
- Different scales
- ...

TU Delft
Delft University of Technology

TNO

Observations/Measurements and uncertainties

Climate, Air and Sustainability

Petroleum Geosciences

Fluid Dynamics

- Representativeness errors
- Unknown reservoir conditi
- Different scales
- …

- Representativeness errors
- Different scales
- …

- Water level
- Inzinking
- Density and velocities
- …

Delft University of Technology

We solve different problems with the same approach (cross-fertilization)

Fluid dynamics

- Estimate essential parameters measuring directly not feasible
- Predict hopper behavior
- Optimize dredging cycle fuel cost, cycle time

Petroleum Geosciences

- Model calibration
  - Integrating information from measurements at different scales
  - Keeping the model evergreen
- Field development plan
  - Optimize production strategies
  - Optimize well locations

Climate, Air and Sustainability

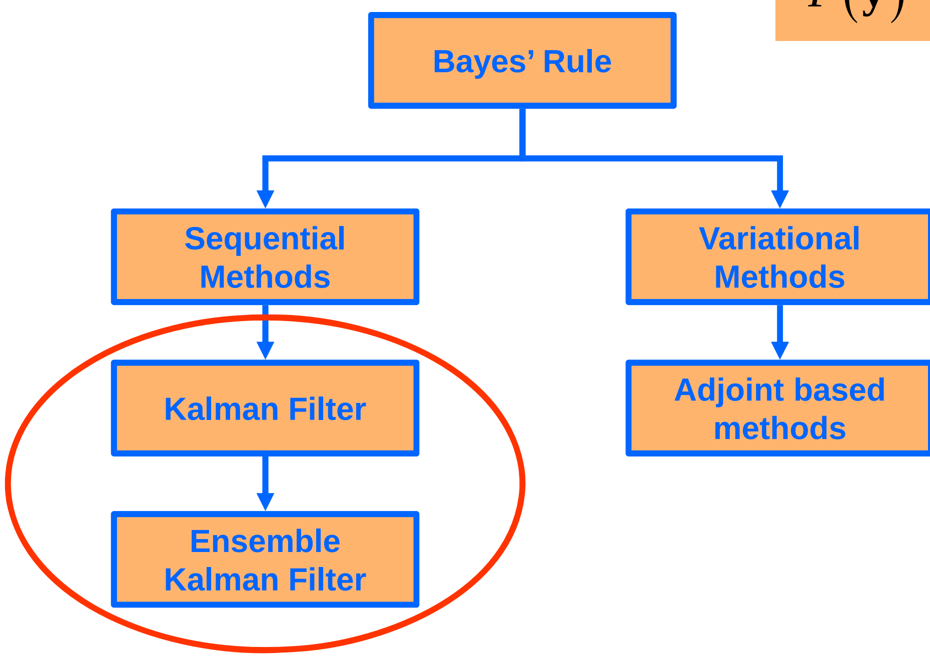
- Optimal estimation of uncertain parameters
  - Optimal estimates for the dynamical parameters
- Predicting peaks of ozone high concentrations
- Reconstructing the emissions sources

Probabilistic Data Assimilation – Bayes' rule

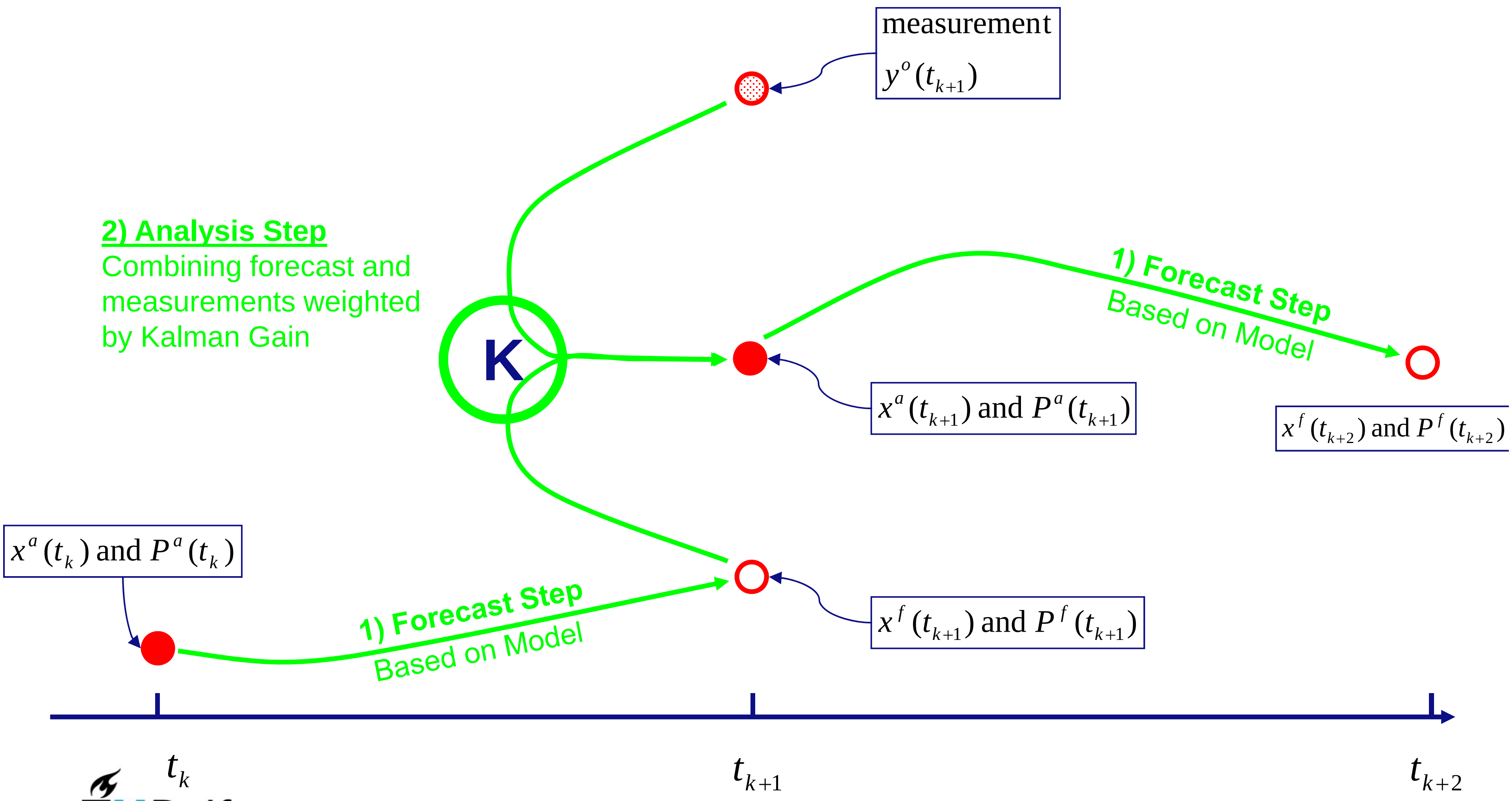
$$P(\mathbf{x}\,|\,\mathbf{y}) = \frac{P(\mathbf{y}\,|\,\mathbf{x})\,P(\mathbf{x})}{P(\mathbf{y})}$$

$$P(\mathbf{x}\,|\,\mathbf{y}) \propto P(\mathbf{y}\,|\,\mathbf{x})\,P(\mathbf{x})$$

$P(\mathbf{x}\,|\,\mathbf{y})$	Posterior probability
$P(\mathbf{x})$	Prior probability
$P(\mathbf{y}\,|\,\mathbf{x})$	Likelihood of observations, given a model
$P(\mathbf{y})$	Probability of observations



Classical Kalman Filter Steps



TU Delft
Delft University of Technology

TNO

System and the measurements:

$$x^t(t_{k+1}) = M(x^t(t_k)) + w(t_k)$$

$$y^o(t_{k+1}) = H(t_{k+1})x^t(t_{k+1}) + v(t_{k+1})$$

$$w \sim N(0,Q)$$

$$v \sim N(0,R)$$

Model and observations

1) Forecast step:

$$x^f(t_{k+1}) = E(x^t(t_{k+1})) = \mathbf{M}(t_k)x^a(t_k)$$

$$P^f(t_{k+1}) = E[(x^t(t_{k+1}) - x^f(t_{k+1}))(x^t(t_{k+1}) - x^f(t_{k+1}))^T]$$

2) Analysis step:

$$x^a(t_{k+1}) = x^f(t_{k+1}) + K(t_{k+1})(y^o(t_{k+1}) - H(t_{k+1})x^f(t_{k+1}))$$

$$P^a(t_{k+1}) = E[(x^t(t_{k+1}) - x^a(t_{k+1}))(x^t(t_{k+1}) - x^a(t_{k+1}))^T]$$

$$K(t_{k+1}) = P^f(t_{k+1})H(t_{k+1})^T[H(t_{k+1})P^f(t_{k+1})H(t_{k+1})^T + R(t_{k+1})]^{-1}$$

Estimation using Kalman Filter

Calculates only the first statistical moments: mean and covariance

Non-classical Kalman Filters

- Classical Kalman Filter assumes:
  - Linearity for the model operator and observation operator.
  - Gaussian distribution for the statistics of the error distribution.

- But in reality, this is usually not the case

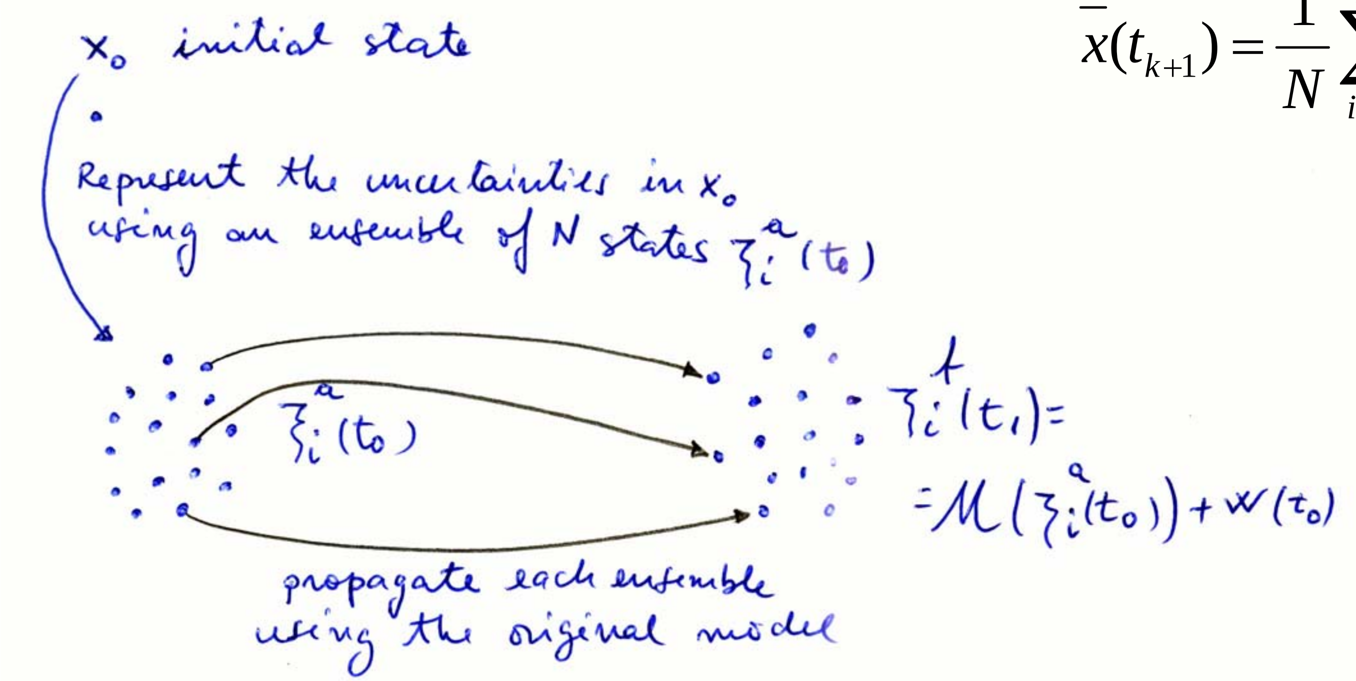
- Remedies:
  - The Extended Kalman filter
    Was used in the Apollo missions, but it is not practical for complex systems because of computational burden.
  - Ensemble Kalman filter and adjoint based methods can be used with a nonlinear model and nonlinear measurement model.

Ensemble Kalman Filter

- Advantages
  1. Can be used for nonlinear models.
  2. Fairly simple to implement.
  3. No need to go into the details of the forward model.
  4. Computational advantages (lower rank covariances)

- Disadvantages
  1. It is very sensitive to the “good” knowledge of the statistics.
  2. Requires a large number of members of the ensemble to converge to the real parameter.

Ensemble Kalman Filter

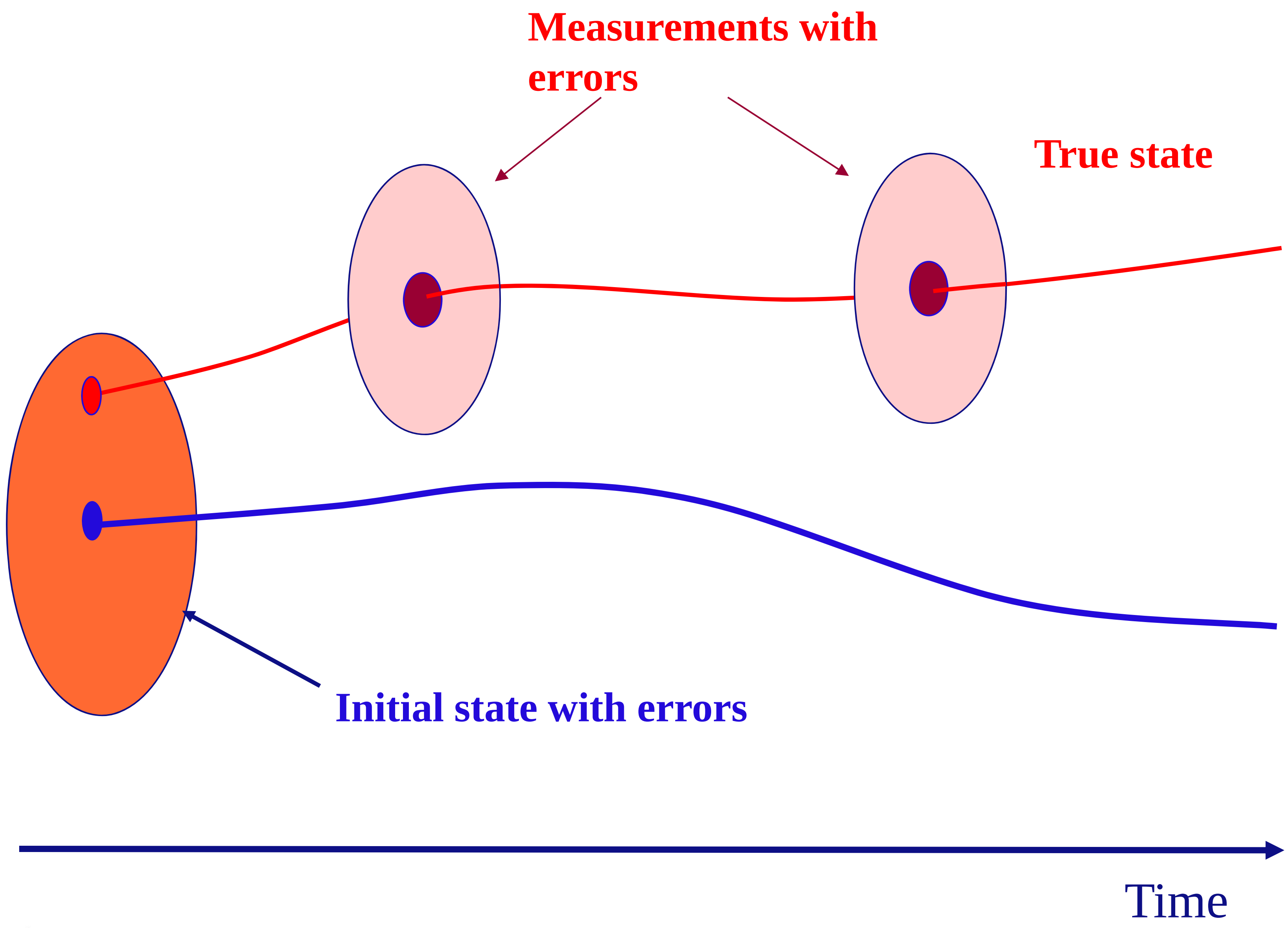


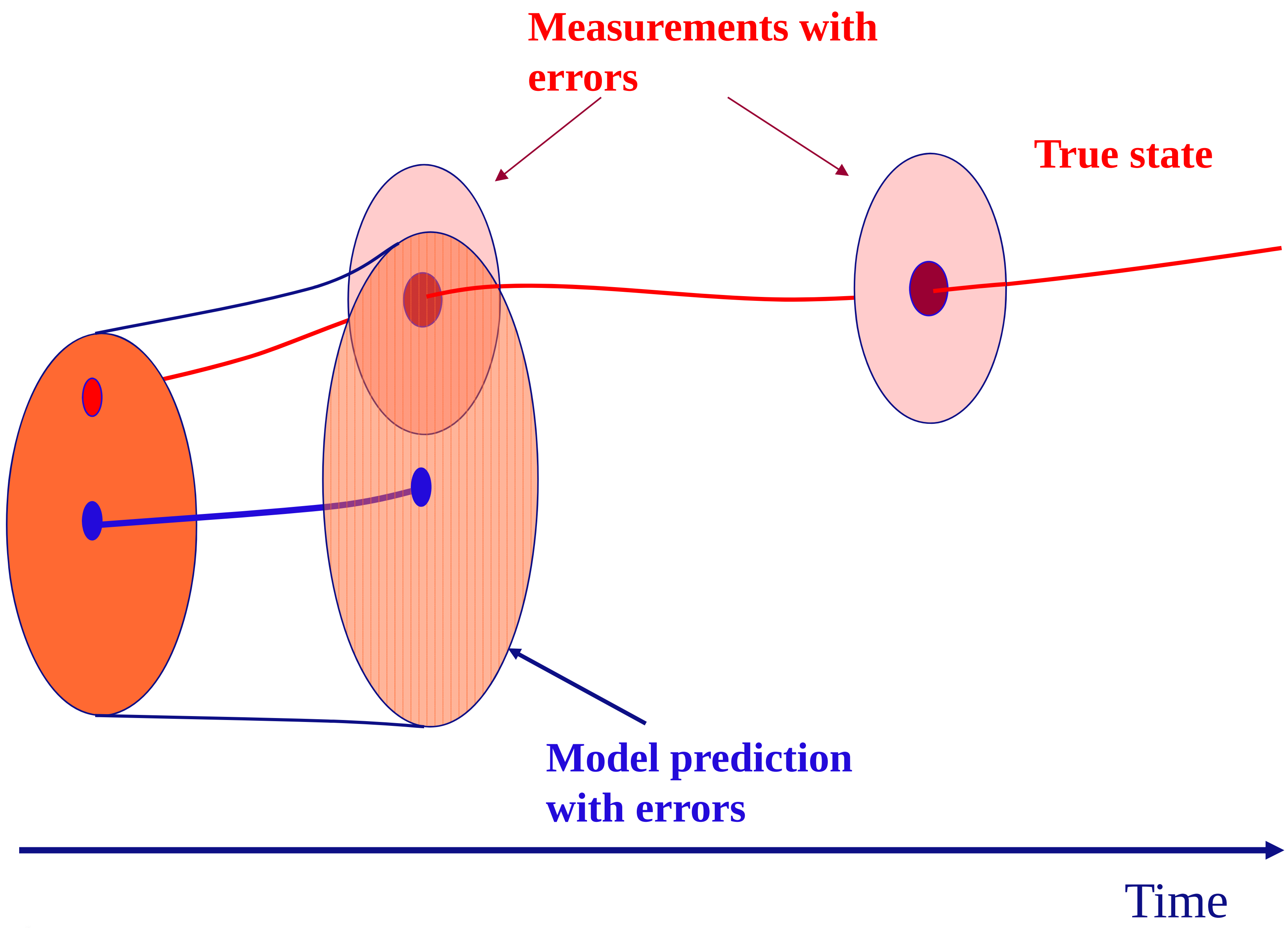
$$\bar{x}(t_{k+1}) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^f(t_{k+1})$$

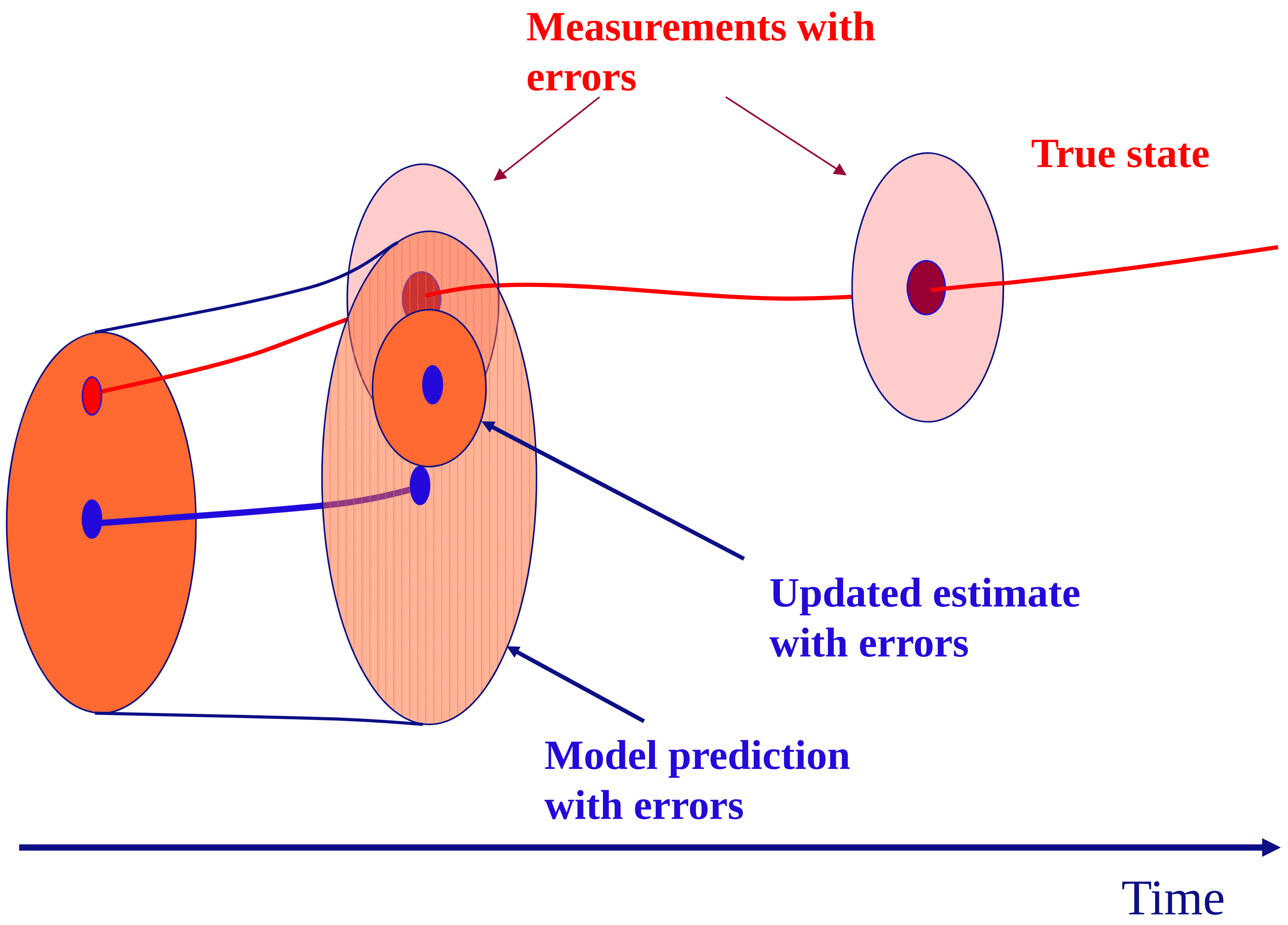
$$\xi_i^a(t_k) = \xi_i^f(t_{k+1}) + K(t_{k+1})[y^o(t_{k+1}) - H(t_{k+1})\xi_i^f(t_{k+1}) + v_i(t_{k+1})]$$

$$P^f(t_{k+1}) \approx P_e^{\,f}(t_{k+1}) = E[(\bar{x}(t_{k+1}) - x^f(t_{k+1}))(\bar{x}(t_{k+1}) - x^f(t_{k+1}))^T]$$

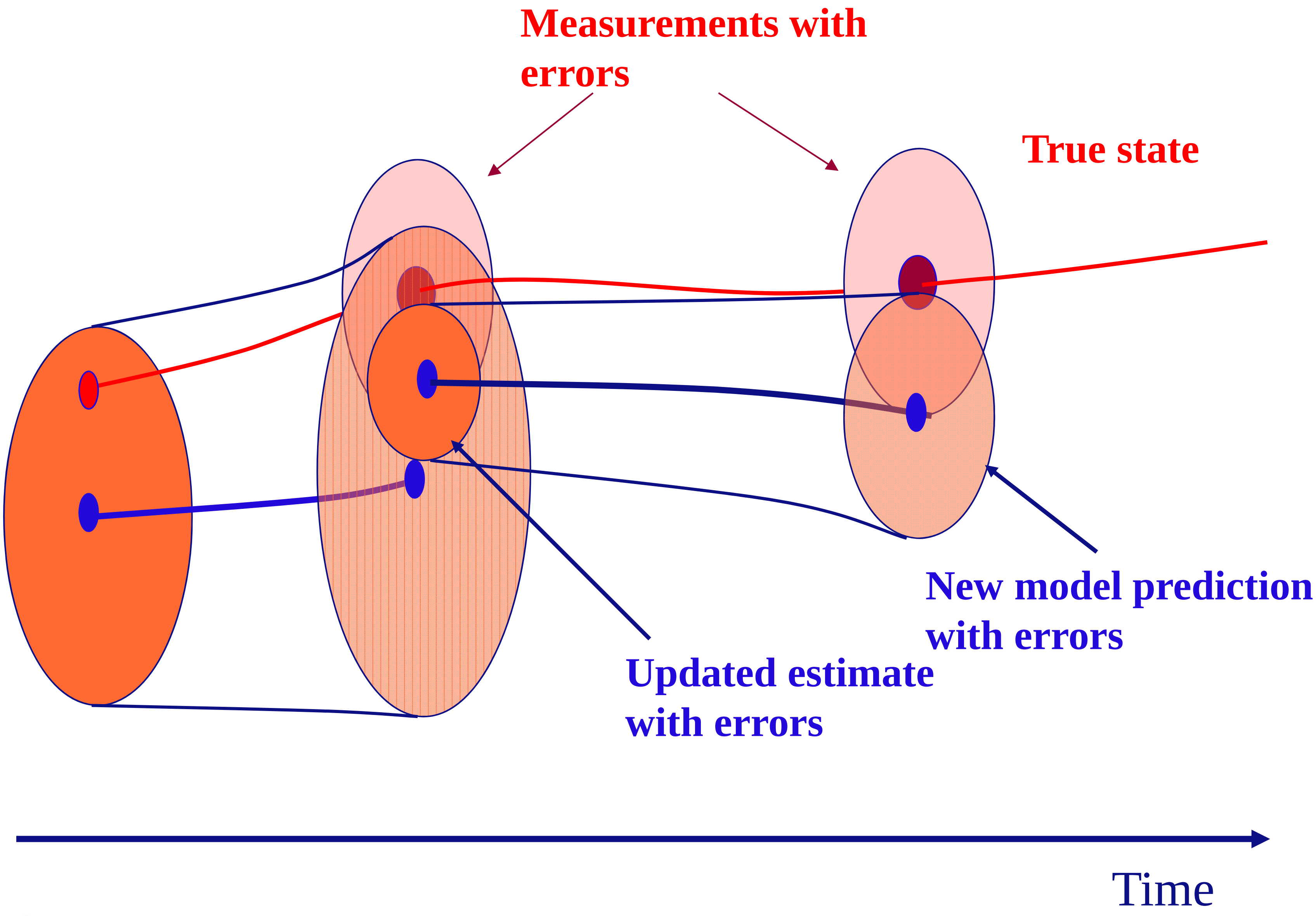
$$P^a(t_{k+1}) \approx P_e^{\,a}(t_{k+1}) = E[(\bar{x}(t_{k+1}) - x^a(t_{k+1}))(\bar{x}(t_{k+1}) - x^a(t_{k+1}))^T]$$

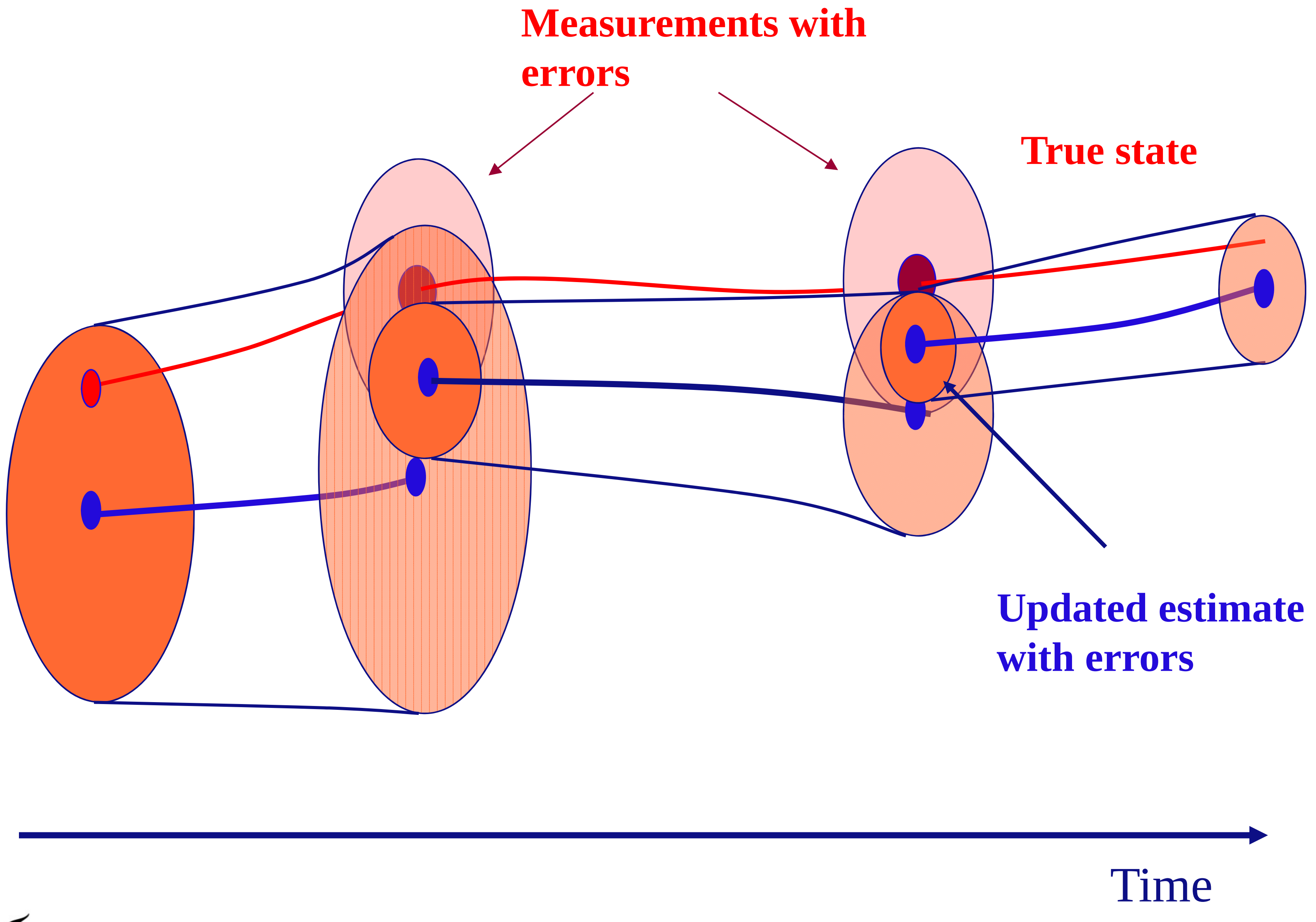

Measurements with errors
True state
Initial state with errors
Time


Measurements with errors
True state
Model prediction with errors
Time

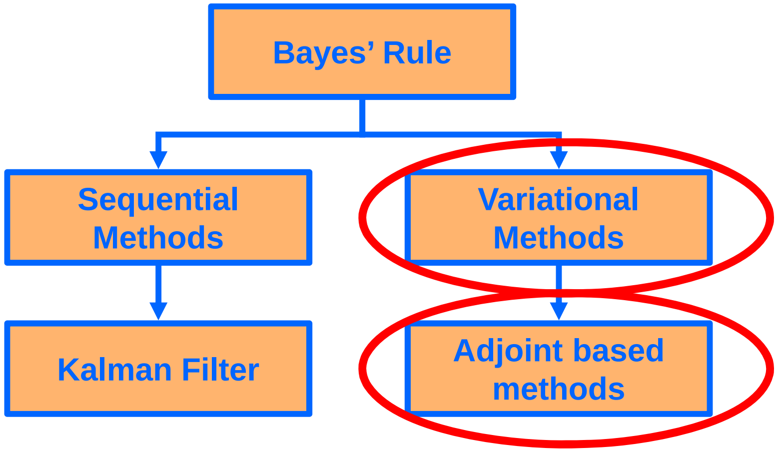


Delft University of Technology


Measurements with errors
True state
New model prediction with errors
Updated estimate with errors
Time


Measurements with errors
True state
Updated estimate with errors
Time

Variational methods



$$\min J(x,u)$$

x represents the <u>state variables</u>,
in our case pressure and saturation

u represents

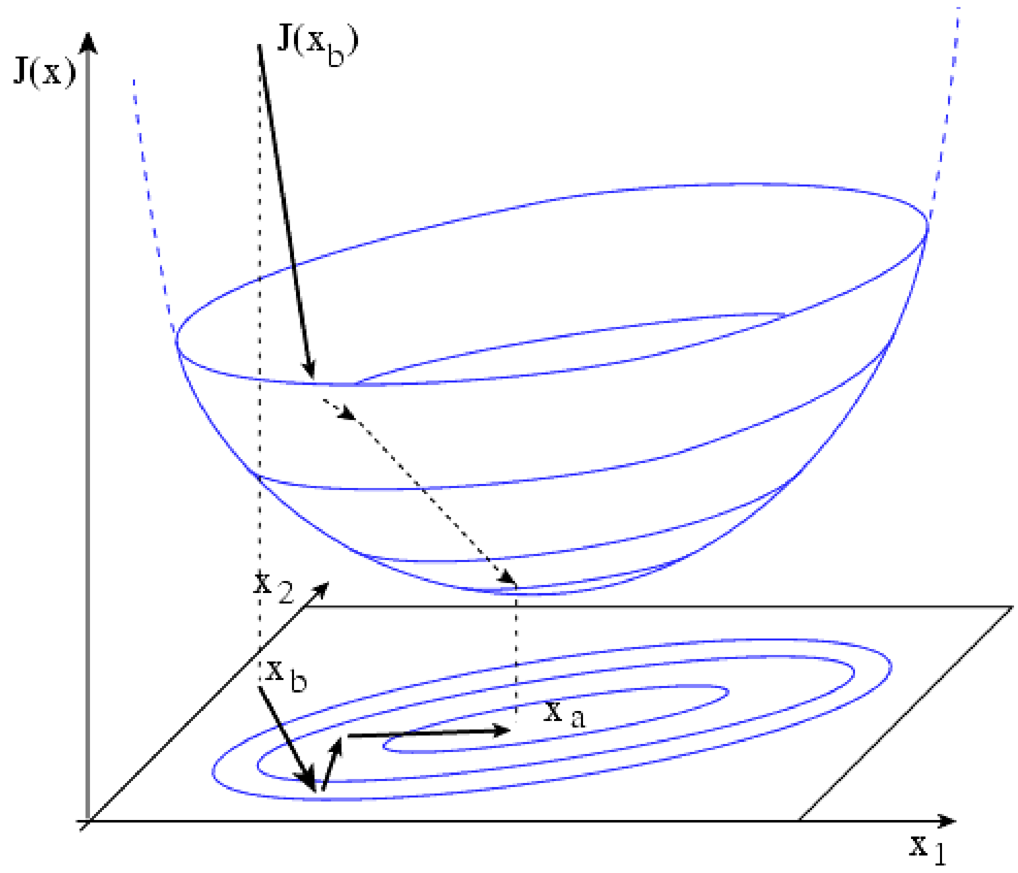
- the <u>reservoir model parameters</u> that we want to estimate in the history matching, or
- the <u>control parameters</u> that we want to optimally set in the field development plan

TU Delft
Delft University of Technology

TNO

Variational methods – the principle

$$\min J(x)$$



Variational methods – the Jacobian, the malefactor

- We need to calculate the gradient (Jacobian)

$$\frac{dJ(x,u)}{du} = \frac{\partial J}{\partial u} + \frac{\partial J}{\partial x}\frac{\partial x}{\partial u} \qquad \left[\frac{dJ(x,u)}{du}\right]^T$$

Derivatives with respect to the parameters

Derivatives with respect to the state variables

adjoint

- u may easily represent 100s of variables, but worse
- x may represent millions of variables, for each time step!

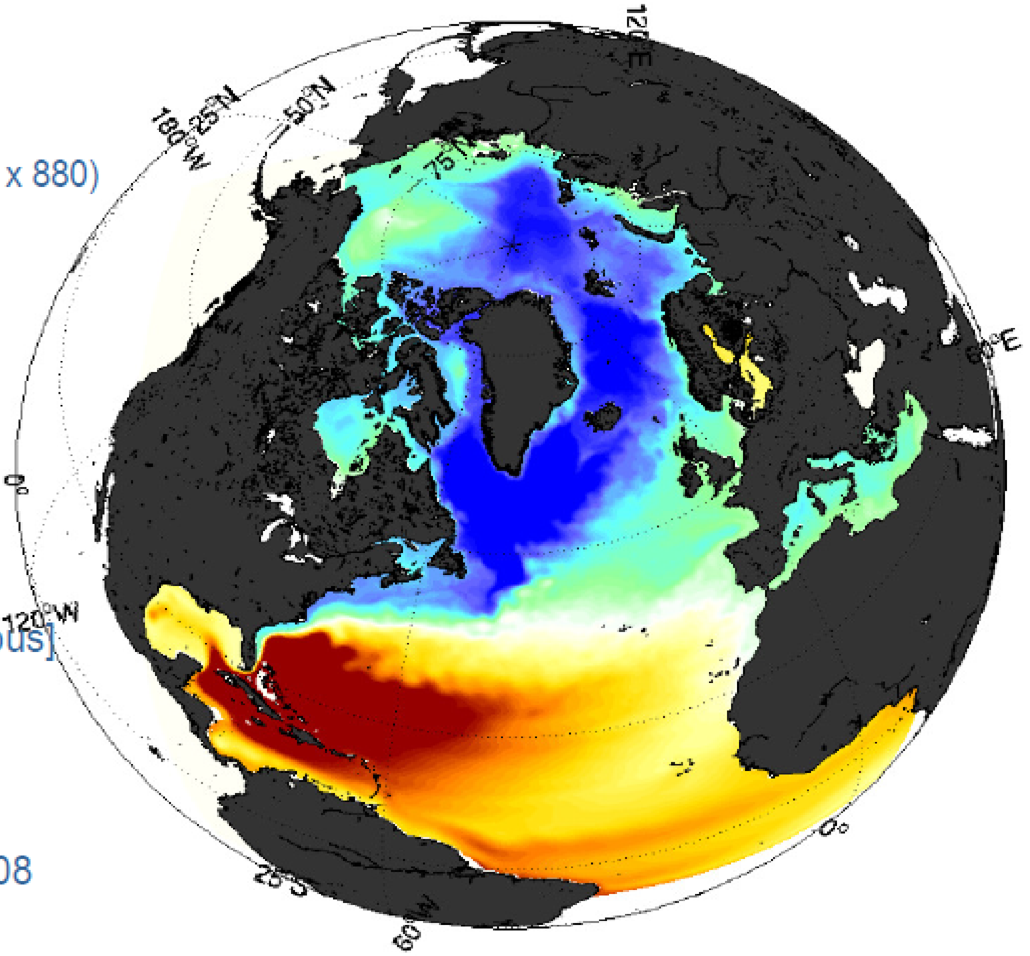
- Options to calculate the Jacobian:
  - Numerical differentiation: computationally not feasible in our case
  - Adjoint method: computationally efficient, but requires significant programming efforts

Challenges

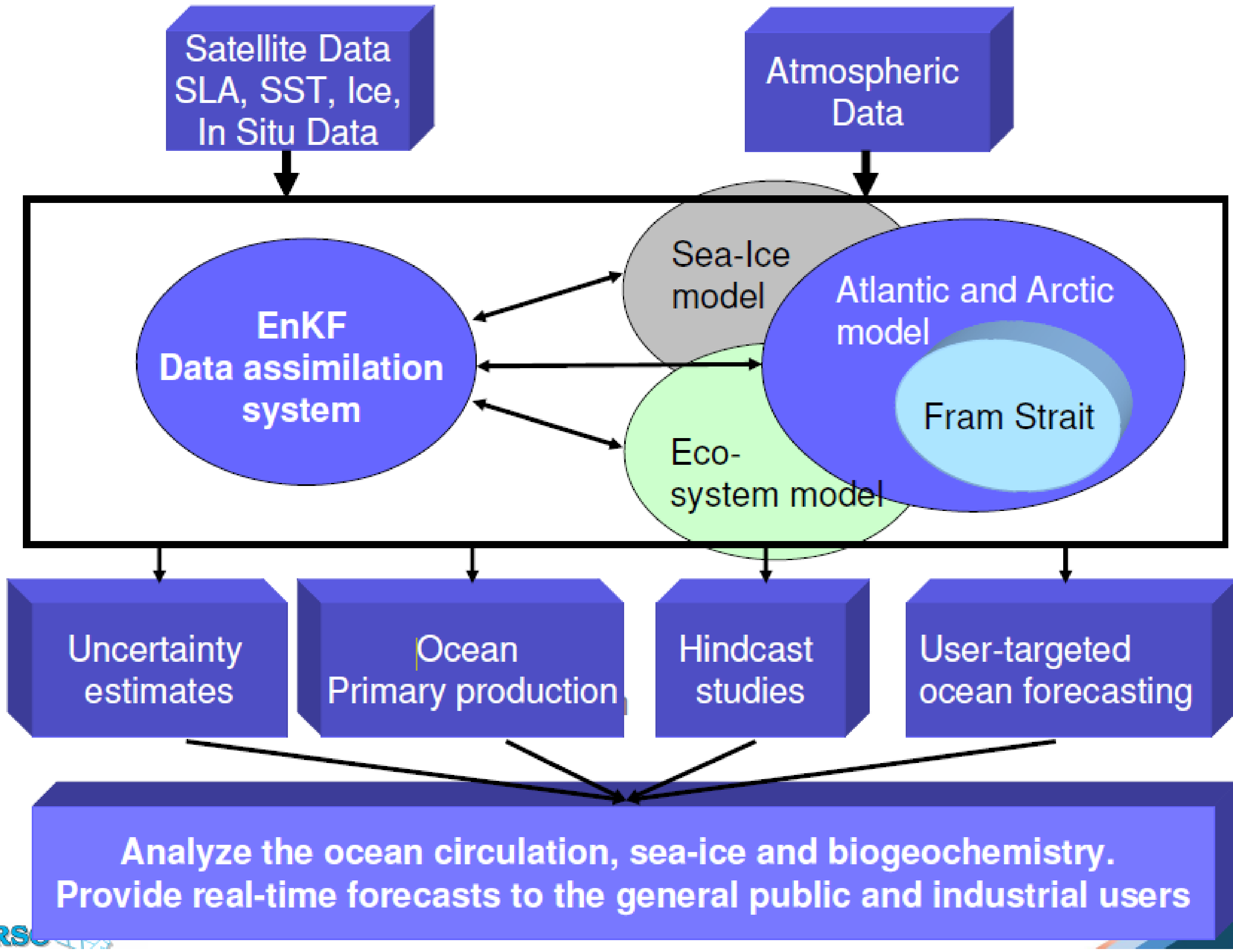
- …

The TOPAZ model system

- TOPAZ3: Atlantic and Arctic
  - HYCOM + EVP sea-ice model
  - 11- 16 km horizontal resolution (800 x 880)
  - 22 hybrid layers
- EnKF
  - 100 members
- Observations
  - Sea Level Anomalies (CLS)
  - Sea Surface Temperatures (NOAA)
  - Sea Ice Concentr. (AMSR, NSIDC)
  - Sea ice drift (CERSAT) [asynchronous]
  - Argo T/S profiles (Coriolis)
- Runs weekly, 10 days forecasts
  - ECMWF forcing
  - Exploited at met.no since March 2008



NERSC


Satellite Data
SLA, SST, Ice,
In Situ Data
Atmospheric
Data
EnKF
Data assimilation
system
Sea-Ice
model
Atlantic and Arctic
model
Fram Strait
Eco-
system model
Uncertainty
estimates
Ocean
Primary production
Hindcast
studies
User-targeted
ocean forecasting
Analyze the ocean circulation, sea-ice and biogeochemistry.
Provide real-time forecasts to the general public and industrial users

Case studies – Highly nonlinear dynamics

2D variables (400 x 600 grid cells)

-Barotropicpressure
-u/v velocity
-ice concentration
-ice thickness

3D variables (400 x 600 x 22 grid cells)

-Temperature
-salinity
-u/v current
-layer thickness

TOTAL: **27.600.000 variables**

Sea level anomalies (satellite, radar altimeters)

-Non linear function of state variables
-100.000 observations every week

Sea-surface temperature (satellite, optical)

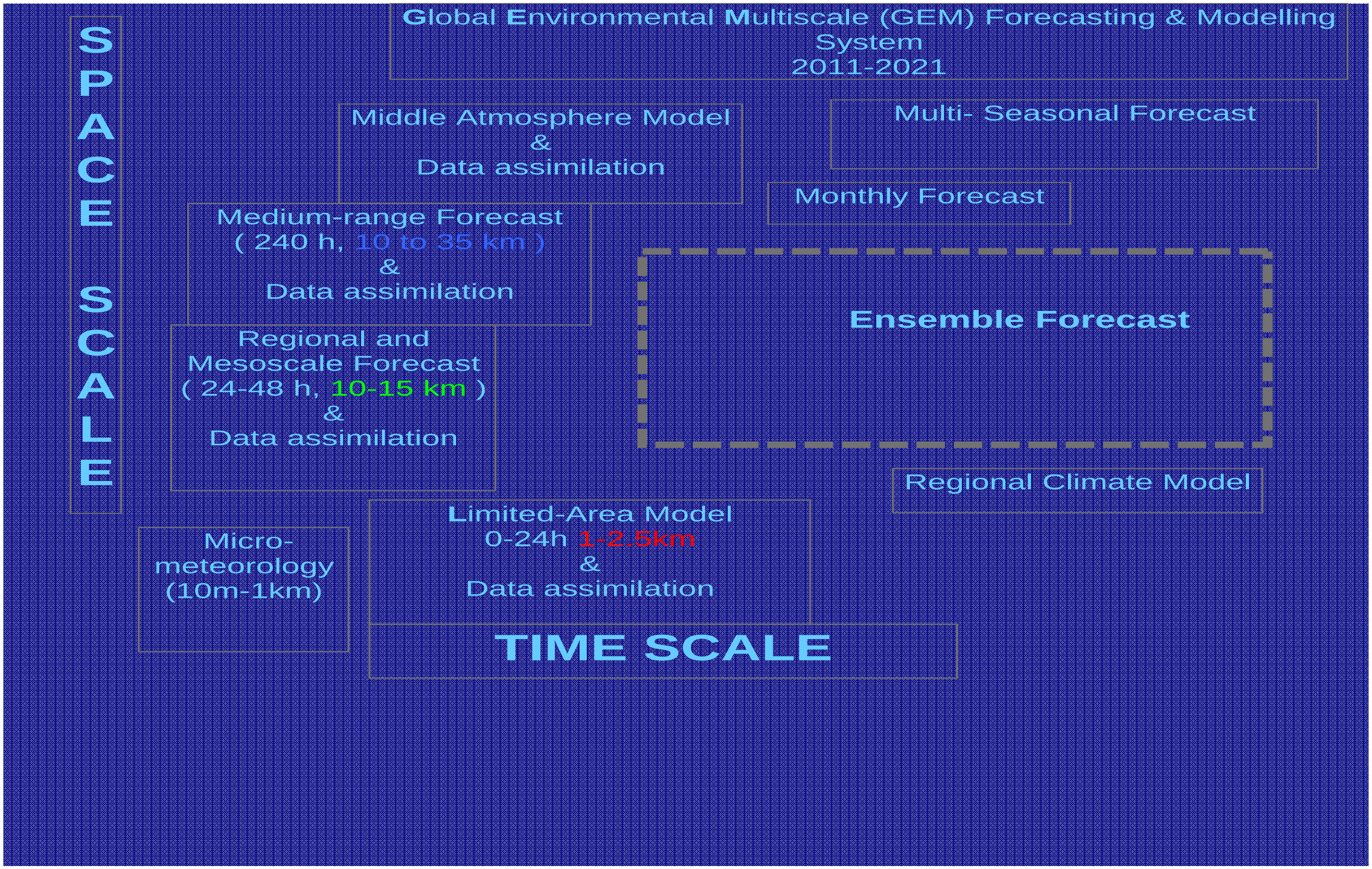
- 8.000 observations every week

Sea-ice concentrations (satellite, microwave)

- 40.000 observations every week

TOTAL: **148.000 measurements**

An unified numerical weather forecasting operational system

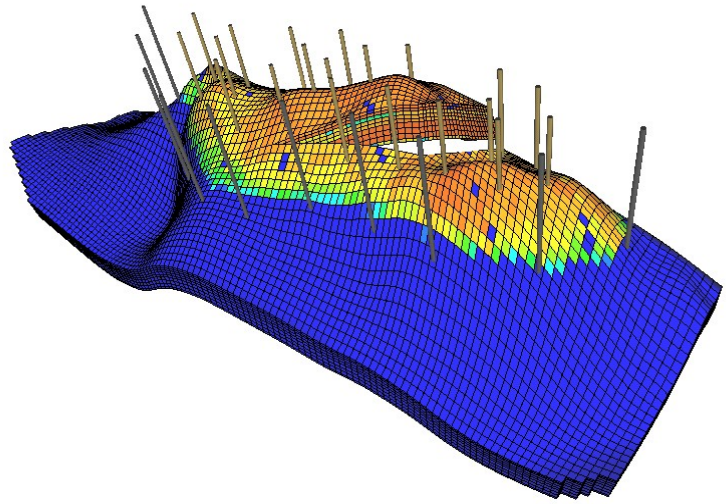


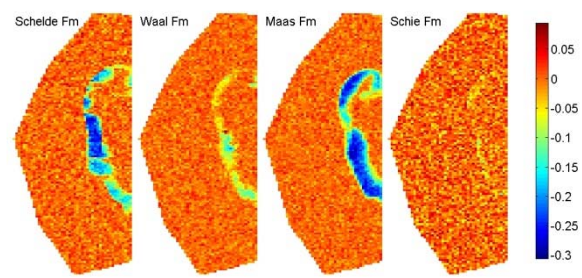
Canadian Meteorological Center, Weather prediction Division

Reservoir management workflow benchmark study

Peters et al., 2010, SPE J.

- Synthetic case
- 44500 active grid cells with 4 values in each grid
- relative perms, initial OWC, vertical transmissibility
- 10 or 20 year production history
- 20 producers and 10 injectors
- time-lapse seismic





Can we optimize the oil production (maximize NPV) over 30 years period?

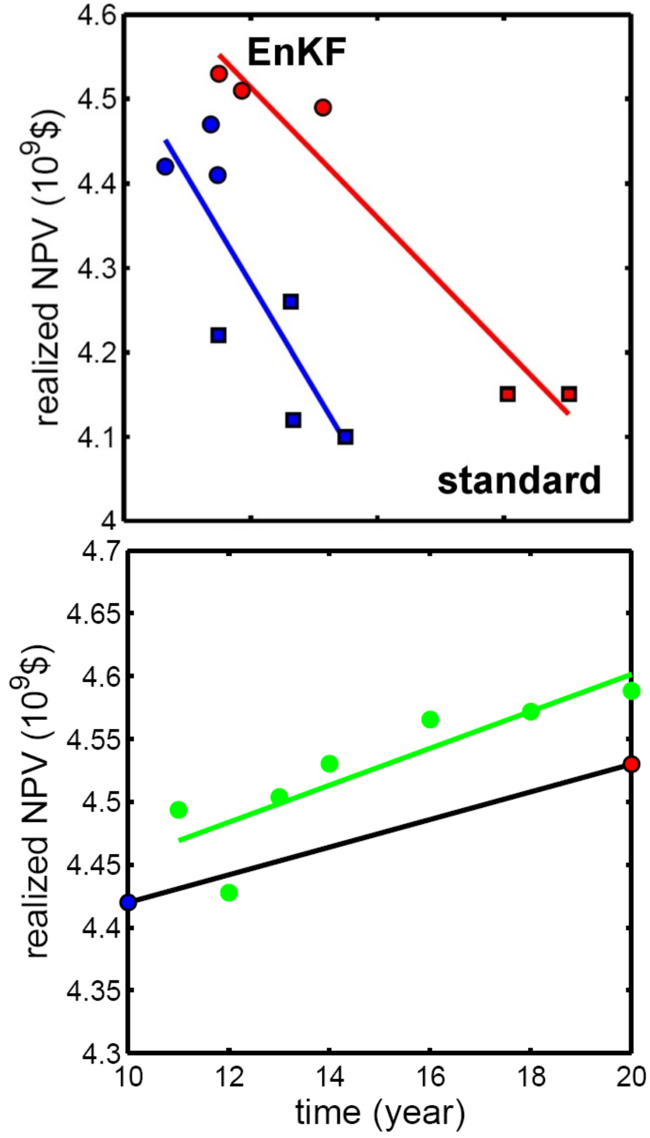
Reservoir management workflow benchmark study

- Estimation of different properties
- Non-linear dynamics
- Two distinct data types
- Different simulators used by the participants

Successes:

- Use of the EnKF as a history matching method was a common factor among the best performers

- Updating models and production strategies more frequently improves the forecast of the final realized NPV.



Seismic History Matching of Fluid Fronts

Trani et al. 2011, submitted to SPE Journal

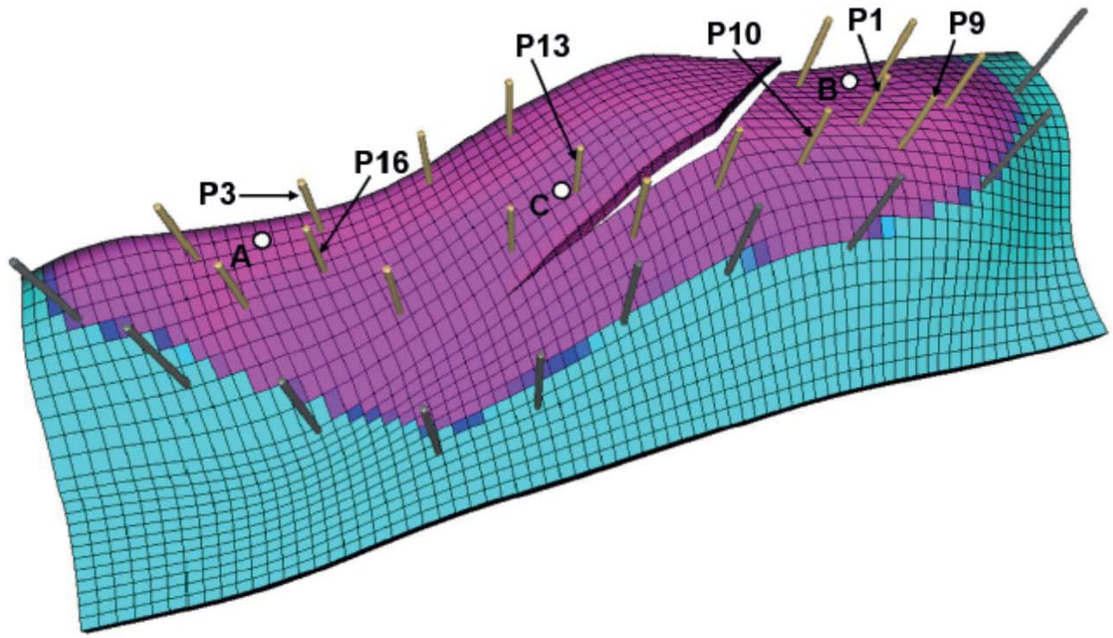
• Synthetic case based on Brugge field

• 20000 active grid cells with 2 values in each grid cell

•14 years of production

• 17 producers and 10 injectors

• a re-parameterization of time-lapse seismic into front arrivals times (no extra inversion required)



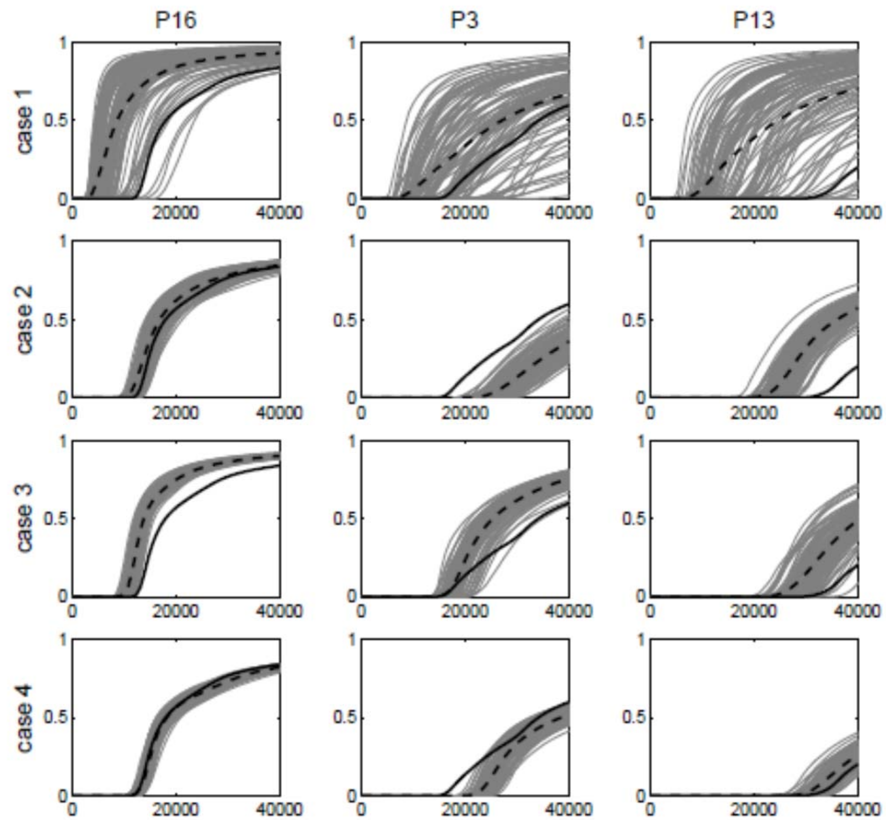
What is the added values of a new parameterization for time –lapse seismic?

Seismic History Matching of Fluid Fronts

- Non-linear dynamics
- Two distinct data types
- MORES simulator

Successes:

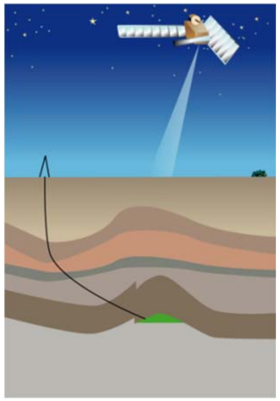
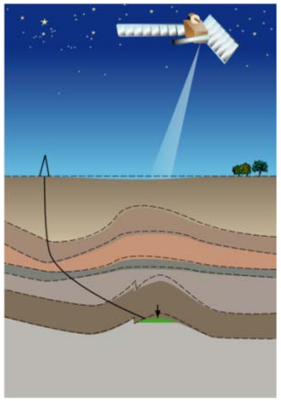
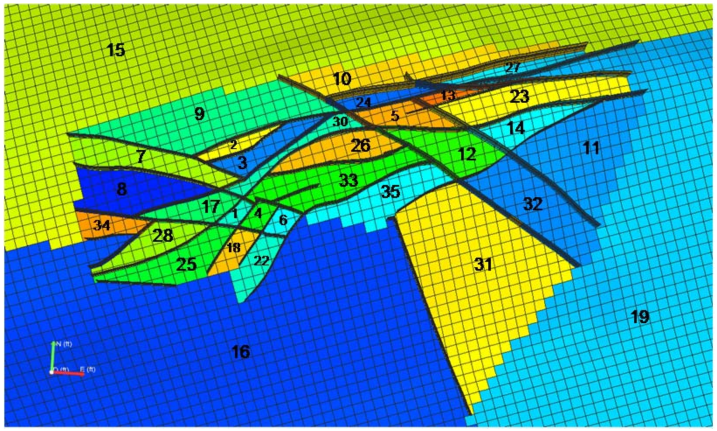
- The new re-parameterization is a success
- No extra inversion required
- Improved match for both production and seismic data => improved forecast skills



Roswinkel Field Case
Joint HM subsidence and well data

Wilschut et al., SPE 141690

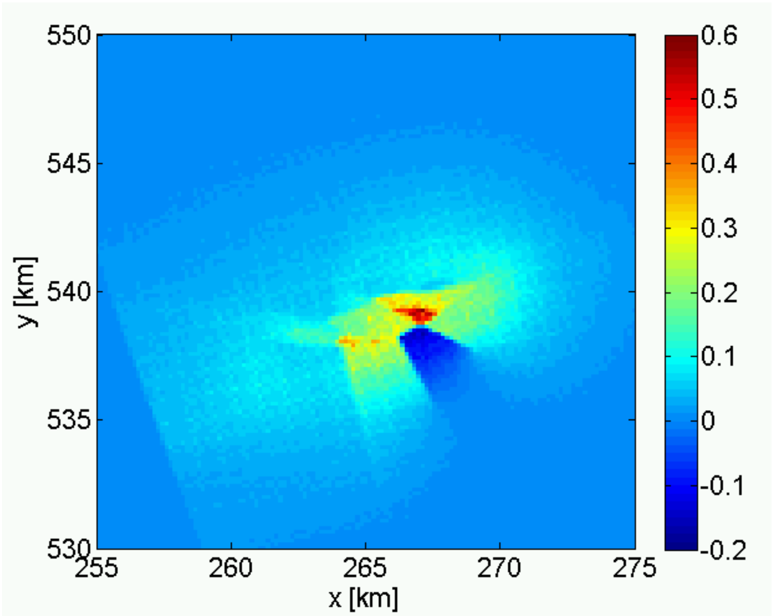
- Heavily faulted gas field in NE-Netherlands
- 35 possible compartments
- GIIP 24.6 bcm
- Production period 1980-2005
- 9 leveling subsidence campaigns
- Max. subsidence 17 cm

Can we identify compartmentalization based on both subsidence and production data?

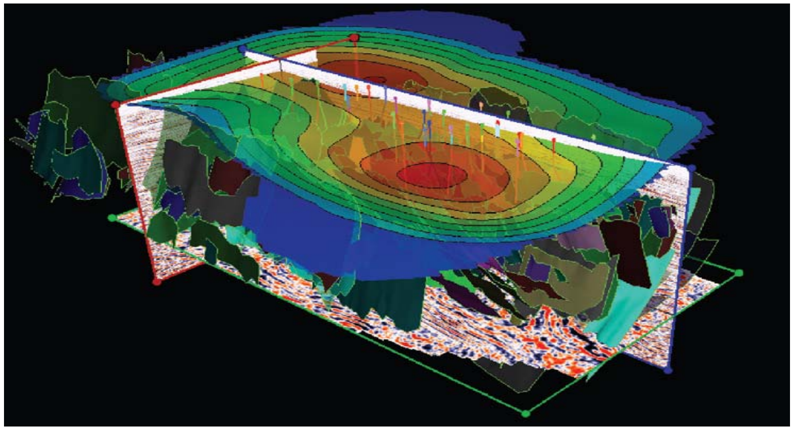
Roswinkel Field Case
Joint HM subsidence and well data

- Estimation of fault properties
- Moderately non-linear
- Two distinct data types
- Simulator: IMEX coupled with geomechanical model

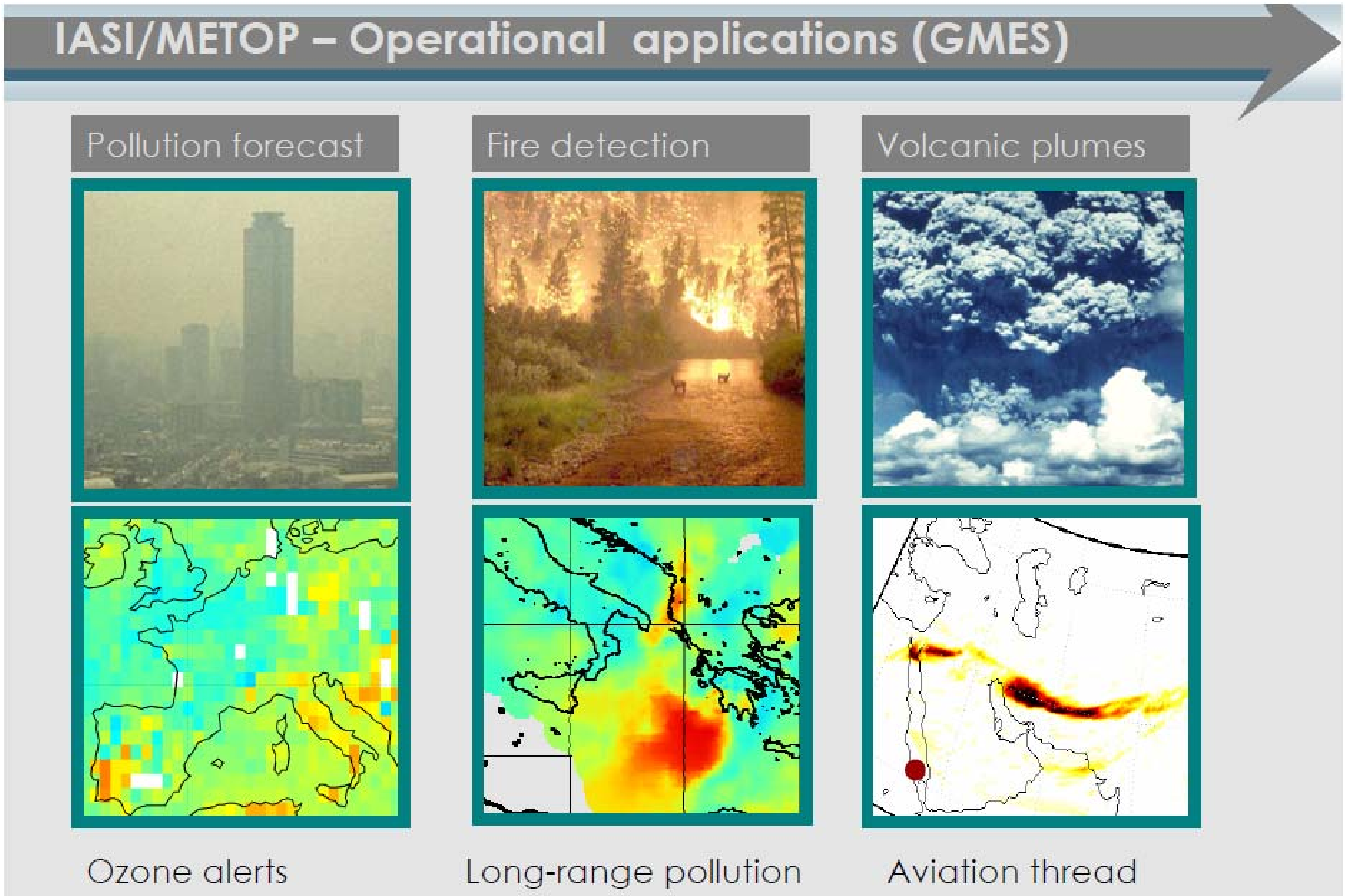


Successes:

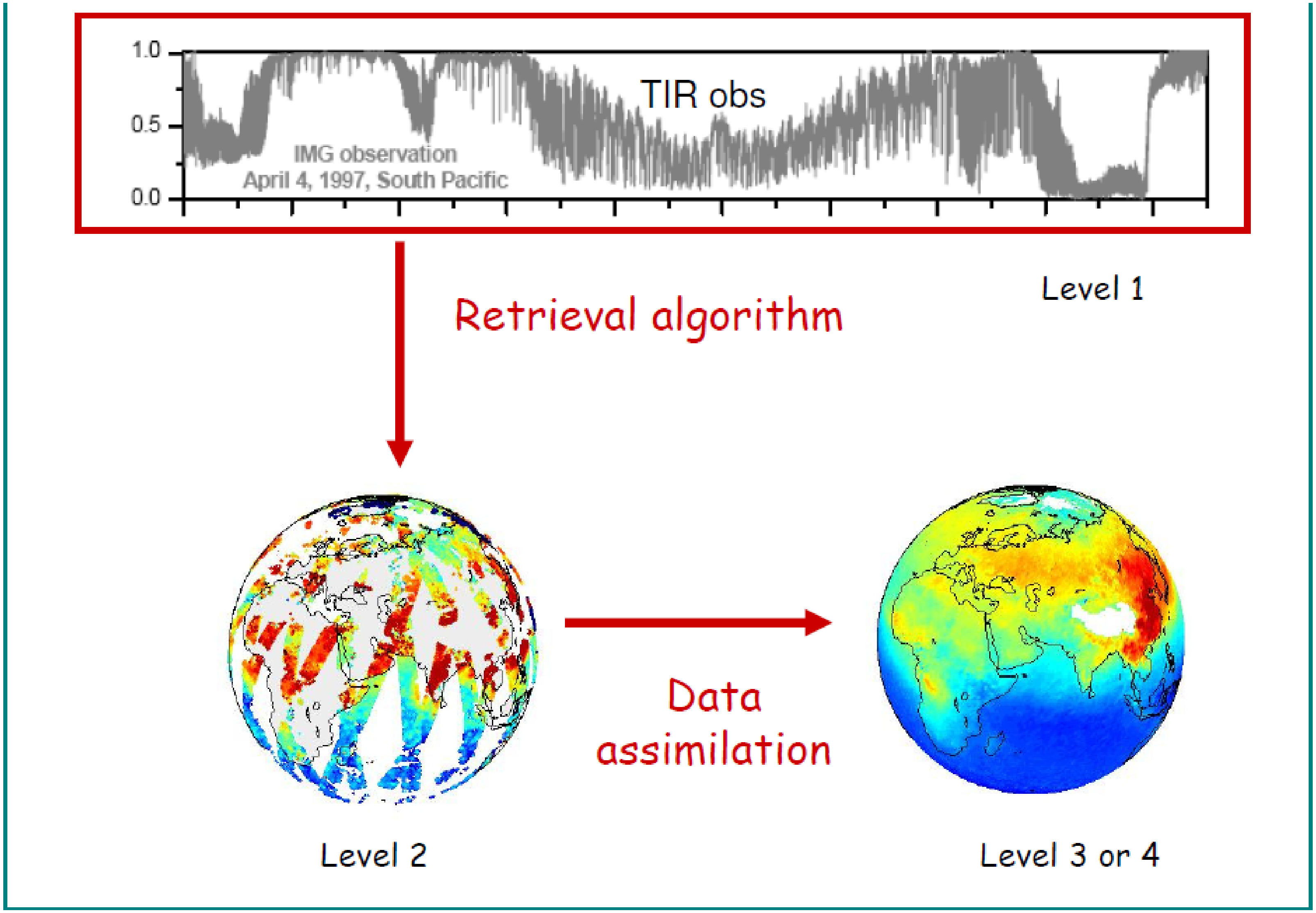
- Added value of second data type
- EnKF can also be used as diagnostic tool



Infrared remote sensing of atmospheric composition and air quality: towards operational applications



TU Delft
Delft University of Technology


1.0
0.5
0.0
TIR obs
IMG observation
April 4, 1997, South Pacific
Level 1
Retrieval algorithm
Level 2
Data assimilation
Level 3 or 4

Conclusions

- Combination between the model and measurements
- Estimation and forecast tool under uncertainties.
- It is very sensitive to the right description of the uncertainties
- Data assimilation is a successful recipe/solution for a lot of different types of applications